

High-Fidelity Audio Compression with Improved RVQGAN

Year: 2023 | Citations: 541 | Authors: Rithesh Kumar, Prem Seetharaman, Alejandro Luebs, I. Kumar, Kundan Kumar

Abstract

Language models have been successfully used to model natural signals, such as images, speech, and music. A key component of these models is a high quality neural compression model that can compress high-dimensional natural signals into lower dimensional discrete tokens. To that end, we introduce a high-fidelity universal neural audio compression algorithm that achieves $\sim 90\times$ compression of 44.1 KHz audio into tokens at just 8kbps bandwidth. We achieve this by combining advances in high-fidelity audio generation with better vector quantization techniques from the image domain, along with improved adversarial and reconstruction losses. We compress all domains (speech, environment, music, etc.) with a single universal model, making it widely applicable to generative modeling of all audio. We compare with competing audio compression algorithms, and find our method outperforms them significantly. We provide thorough ablations for every design choice, as well as open-source code and trained model weights. We hope our work can lay the foundation for the next generation of high-fidelity audio modeling.